

Vertex Percolation on Expander Graphs

Sonny Ben-Shimon *

Michael Krivelevich[†]

July 1, 2008

Abstract

We say that a graph $G = (V, E)$ on n vertices is a β -expander for some constant $\beta > 0$ if every $U \subseteq V$ of cardinality $|U| \leq \frac{n}{2}$ satisfies $|N_G(U)| \geq \beta|U|$ where $N_G(U)$ denotes the neighborhood of U . In this work we explore the process of deleting vertices of a β -expander independently at random with probability $n^{-\alpha}$ for some constant $\alpha > 0$, and study the properties of the resulting graph. Our main result states that as n tends to infinity, the deletion process performed on a β -expander graph of bounded degree will result with high probability in a graph composed of a giant component containing $n - o(n)$ vertices that is in itself an expander graph, and constant size components. We proceed by applying the main result to expander graphs with a positive spectral gap. In the particular case of (n, d, λ) -graphs, that are such expanders, we compute the values of α , under additional constraints on the graph, for which with high probability the resulting graph will stay connected, or will be composed of a giant component and isolated vertices. As a graph sampled from the uniform probability space of d -regular graphs with high probability is an expander and meets the additional constraints, this result strengthens a recent result due to Greenhill, Holt and Wormald about vertex percolation on random d -regular graphs. We conclude by showing that performing the above described deletion process on graphs that expand sub-linear sets by an unbounded expansion ratio, with high probability results in a connected expander graph.

1 Introduction

In this paper we analyze the process of deleting vertices independently at random from an expander graph and describe typical properties and the structure of the resulting graph. We focus on the case where the initial graph, G , is of bounded degree and the deletion probability equals $n^{-\alpha}$, for any fixed $\alpha > 0$, where n denotes the number of vertices in G . We are mainly interested in investigating when the resulting graph with high probability will possess some expansion properties as will be discussed in Section 1.3. In a recent paper of Greenhill, Holt and Wormald [9], the authors perform a very similar analysis where the initial graph is sampled from the uniform probability space of all d -regular graphs for some fixed $d \geq 3$. Our current result, generalizing and improving [9], can be interpreted as providing sufficient deterministic conditions on

*School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: sonny@post.tau.ac.il. Research conducted as part of the author's Ph.D. thesis under the supervision of Prof. Michael Krivelevich.

[†]School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2006-322, and by grant 526/05 from the Israel Science Foundation.

the initial graph that imply the result of [9]. We are also able to prove some results when the initial graph has an unbounded expansion ratio, and apply it to the case of random d -regular graphs when $d = o(\sqrt{n})$.

1.1 Notation

Given a graph $G = (V, E)$, the *neighborhood* $N_G(U)$ of a subset $U \subseteq V$ of vertices is the set of vertices defined by $N_G(U) = \{v \notin U : v \text{ has a neighbor in } U\}$. For any $f : [\lfloor \frac{n}{2} \rfloor] \rightarrow \mathbb{R}^+$, we say that a graph $G = (V, E)$ on n vertices is an f -*expander* if every $U \subseteq V$ of cardinality $|U| \leq \frac{n}{2}$ satisfies: $|N_G(U)| \geq f(|U|) \cdot |U|$. When f is a constant function equal to some $\beta > 0$ we say that G is a β -*expander*. When a function $f : A \rightarrow \mathbb{R}^+$ satisfies: $f(a) \geq c$ for any $a \in A$, where $c \geq 0$ is a constant, we simply write $f \geq c$.

Expanders in general are highly-connected sparse graphs. There are many other notions and definitions of expander graphs different from the one described above, some of which will be addressed in the coming sections. Expander graphs is a subject of utmost importance to the fields of both applied and theoretical Computer Science, Combinatorics, Probability Theory etc. Monograph [11] provides an excellent survey on expander graphs and their applications.

In our setting, we start with a graph $G = (V, E)$ on n vertices. We delete every vertex of V with probability $p = n^{-\alpha}$ for some fixed $\alpha > 0$ independently at random. To simplify notation, from here on, we will denote the resulting graph of this process by $\hat{G} = (\hat{V}, \hat{E})$, and for every $X \subseteq V$, we denote by $\hat{X} = X \cap \hat{V}$ the subset of X that was not deleted by the deletion process. We denote by \hat{n} the cardinality of \hat{V} , by R the set of deleted vertices, i.e. $\hat{G} = G[V \setminus R]$, and by r its cardinality, i.e. $\hat{n} = n - r$. When considering the neighborhood in the graph \hat{G} of a subset of vertices $U \subseteq \hat{V}$ we denote it by $N_{\hat{G}}(U)$.

The main research interest of this paper is the asymptotic behavior of properties of the graph \hat{G} as we let the number of vertices, n , grow to infinity. In this context, one needs to be precise when formulating such claims. When stating an asymptotic claim for every graph G on n vertices that satisfies a set of properties \mathcal{P}_n (the properties may depend on n), one actually means that for every family of graphs $\mathcal{G} = \{G_n\}$, such that G_n is a graph on n vertices satisfying \mathcal{P}_n , there exists a value n_0 such that the claim is correct for every G_n where $n > n_0$. We say that an event \mathcal{A} in our probability space occurs with high probability (or w.h.p. for brevity) if $\Pr[\mathcal{A}] \rightarrow 1$ as n goes to infinity. Therefore, from now on and throughout the rest of this work, we will always assume n to be large enough. We use the usual asymptotic notation. That is, for two functions of n , $f(n)$ and $g(n)$, we denote $f = O(g)$ if there exists a constant $C > 0$ such that $f(n) \leq C \cdot g(n)$ for large enough values of n ; $f = o(g)$ or $f \ll g$ if $f/g \rightarrow 0$ as n goes to infinity; $f = \Omega(g)$ if $g = O(f)$; $f = \omega(g)$ or $f \gg g$ if $g = o(f)$; $f = \Theta(g)$ if both $f = O(g)$ and $f = \Omega(g)$.

1.2 Motivation

Let $\mathcal{G}_{n,d}$ denote the random graph model consisting of the uniform distribution of all d -regular graphs on n labeled vertices (where dn is even). One of the motivations of this paper is the following result, recently proved by Greenhill, Holt and Wormald in [9].

Theorem 1.1 (Greenhill, Holt and Wormald [9]). *For every fixed $\alpha > 0$ and fixed $d \geq 3$ there exists a constant $\beta > 0$, such that if $p = n^{-\alpha}$ and G is a graph sampled from $\mathcal{G}_{n,d}$, then w.h.p. \hat{G} has a connected component of size $n - o(n)$ that is a β -expander and all other components are of bounded size. Moreover,*

1. if $\alpha > \frac{1}{2(d-1)}$, w.h.p. all small connected components of \widehat{G} are isolated vertices.
2. if $\alpha \geq \frac{1}{d-1}$, w.h.p. \widehat{G} is connected.

Theorem 1.1 suggests a few questions that may be of interest to address. First, one might consider the question whether the deletion probability p for which the desired properties hold is best possible. This question has been answered in [9]. Simple probabilistic arguments show that the above result is indeed optimal in the sense that if we let $\alpha = o(1)$, the largest component of \widehat{G} will contain w.h.p. many induced paths of length $O(1/\alpha)$, and hence cannot be an expander. Next, one may ask what are the properties of random d -regular graphs that make the above claim true. One of the research motivations of this paper is precisely that, as will be described below. Moreover, Item 2 of Theorem 1.1 does not seem to be optimal due to the following argument. As random d -regular graphs (for constant values of d) w.h.p. locally look like trees (i.e. there are very few cycles of constant length) it would seem natural to think that to disconnect such a graph one would need to find the deletion probability that is “just enough” to disconnect a single vertex. A simple first moment argument would imply that $\alpha > \frac{1}{d}$ should suffice. In Section 3.2 we confirm this hypothesis in the more general setting of pseudo-random (n, d, λ) -graphs. Lastly, Theorem 1.1 does not consider the case of sampling a random d -regular graph when $d = \omega(1)$, i.e. d goes to infinity with n . This setting is addressed in Section 4.

1.3 Main result

The main result of this paper states that the deletion of vertices of an expander graph G independently at random with probability $n^{-\alpha}$ w.h.p. results in \widehat{G} containing one giant components that is in itself an expander graph. Moreover, the expansion properties of G imply a bound on the sizes of the small connected components of \widehat{G} .

Theorem 1.2. *For every fixed $\alpha, c > 0$ and fixed $\Delta > 0$, there exists a constant $\beta > 0$, such that if G is an f -expander graph on n vertices of bounded maximum degree Δ , and $f \geq c$, then w.h.p. \widehat{G} has a connected component of size $n - o(n)$ that is a β -expander, and the rest of its connected components have at most $K - 1$ vertices, where*

$$K = \min \left\{ u : \forall k \geq u \quad kf(k) > \frac{1}{\alpha} \right\}. \quad (1)$$

We note that K is well defined as $f \geq c$ implies that $K < \frac{1}{c\alpha}$.

As mentioned in Section 1.2, Theorem 1.2 is optimal with respect to the deletion probability if we require the giant component of the resulting graph to possess expansion properties.

It is well known that for fixed $d \geq 3$ random d -regular graphs are w.h.p. expander graphs. Thus our result strengthens Theorem 1.1, as will be formalized in Section 3.3. It should be stressed that the techniques used in the present paper and in [9] are quite different. Whereas in [9] the analysis is done directly in the so called Configuration Model in a probabilistic setting, we rely upon a deterministic property of a graph, namely, being an expander. The approach of first proving some claim under deterministic assumptions, and then showing that these conditions appear w.h.p. in some probability space, allows us to, arguably, simplify the proof, and to get a strengthened result for families of pseudo-random graphs and the random d -regular graph.

1.4 Related work

The process of random deletion of vertices of a graph received rather limited attention, mainly in the context of faulty storage (see e.g. [2]), communication networks, and distributed computing. For instance, the main motivation of [9] is the SWAN peer-to-peer [10] network whose topology possess some properties of d -regular graphs, and may have faulty nodes. Other works are mainly interested in connectivity and routing in the resulting graph after performing (possibly adversarial) vertex deletions on some prescribed graph topologies.

The process of deleting edges, sometimes referred to by *edge-percolation* (or *bond-percolation*) has been more extensively studied. The main interest of edge-percolation is the existence of a “giant component”, i.e. a connected component consisting of a linear size of the vertices, in the resulting graph. When the initial graph is taken to be K_n , edge-percolation becomes the famous $\mathcal{G}(n, p)$ random graph model. In [1, 8, 14] the edge percolation on an expander graph is considered, the authors determine the threshold of the deletion probability at which the giant component emerges w.h.p.. It should be noted that in the context of this paper the expected number of deleted vertices is far lower than permissible in order to retain a giant component in the graph, as is clearly seen in Lemma 2.2.

1.5 Organization of paper

The rest of the paper is organized as follows. In Section 2 we give a proof of Theorem 1.2. We proceed in Section 3 to a straightforward application of our result to expander graphs arising from constraints on the spectrum of the graph. We continue in Section 3.2 to the particular case of (n, d, λ) -graphs, and under additional constraints on the graph compute the values of α for which the resulting graph will w.h.p. stay connected or will be composed of a giant component and isolated vertices. In Section 3.3 we show that a graph sampled from the uniform probability space of d -regular graphs satisfies all constraints, providing an alternative proof of the main result of [9] and even improving it. As a final result, in Section 4 we analyze the case of graphs of unbounded expansion ratio for sub-linear sets with the same deletion probability, and extend our result to random d -regular graphs where $1 \ll d \ll \sqrt{n}$. We conclude in Section 5 with a short summary and open problems for further research.

2 Proof of Theorem 1.2

Let G be an f -expander, where $f \geq c$ for some constant $c > 0$. The number of deleted vertices, r , is clearly distributed by $r \sim \mathcal{B}(n, p)$, and hence by the Chernoff bound (see e.g. [3]) r is highly concentrated around its expectation.

Claim 2.1. *W.h.p. $(1 - o(1))n^{1-\alpha} \leq r \leq (1 + o(1))n^{1-\alpha}$.*

As for $\alpha > 1$ w.h.p. no vertices are deleted from the graph G and the proof of Theorem 1.2 becomes trivial, we will assume from now on that $\alpha \leq 1$. Denote by $\hat{V}_1, \dots, \hat{V}_s$ the partition of \hat{V} to its connected components ordered in descending order of cardinality. We call \hat{V}_1 the *big* component of \hat{G} , and $\hat{V}_2, \dots, \hat{V}_s$ the *small* components of \hat{G} .

Lemma 2.2. *W.h.p. $|\hat{V}_1| \geq (1 - Cn^{-\alpha})n$ for any $C > \frac{1+c}{c}$.*

Proof. First, we show that $|\widehat{V}_1| > \frac{n}{2}$. Assume otherwise, and take $\widehat{W} = \bigcup_{i=1}^j \widehat{V}_i$ for some $j \in [s]$ such that $\frac{n}{4} \leq |\widehat{W}| \leq \frac{n}{2}$. Such a j surely exists. By our condition on f , we have that $|N_G(\widehat{W})| \geq c|\widehat{W}| = \Theta(n)$. But surely, $N_G(\widehat{W}) \subseteq R$, and hence, by Claim 2.1 $|N_G(\widehat{W})| = o(n)$, a contradiction. Now, set $\widehat{U} = \widehat{V} \setminus \widehat{V}_1$. From the above, it follows that $|\widehat{U}| < \frac{n}{2}$. Clearly, $N_G(\widehat{U}) \subseteq R$, and $|N_G(\widehat{U})| \geq c|\widehat{U}|$. Putting these together yields that $|\widehat{U}| \leq \frac{|R|}{c}$, and hence, by Claim 2.1, w.h.p. $|V \setminus \widehat{V}_1| = |R \cup \widehat{U}| \leq (1 + o(1)) \frac{1+c}{c} n^{1-\alpha}$, completing the proof. \square

In a graph H , we call a subset of vertices $U \subseteq V(H)$ *connected* if the corresponding spanning subgraph $H[U]$ is connected. The following well known lemma (see e.g. [13, Exercise 11, p.396]) helps us to bound the number of connected subsets of vertices in a graph of bounded maximum degree.

Lemma 2.3. *If $H = (V, E)$ is a graph of maximum degree D , then V contains at most $\frac{|V|(De)^k}{k}$ connected subsets of cardinality k .*

Keeping in mind that Δ is a constant, Lemma 2.3 turns out to be quite crucial to our forthcoming calculations, for it allows us to bound probability of events using union bound arguments by summing over connected subgraphs of a prescribed cardinality instead of summing over all subgraphs of the respective cardinality. We continue by showing that w.h.p. all small connected components of \widehat{G} must span less than K vertices.

Lemma 2.4. *W.h.p. every small connected component of \widehat{G} is of cardinality at most $K - 1$.*

Proof. By Lemma 2.2, every small connected component of \widehat{G} will be of cardinality at most $\bar{u} = O(n^{1-\alpha})$. Let $U \subseteq V$ be a subset of vertices of cardinality $u \leq \bar{u}$, and let $W = N_G(U)$ be its neighborhood in G , where $|W| = w$. We first bound the probability that \widehat{U} is a small connected component of \widehat{G} by the probability that all of W was deleted,

$$\Pr \left[\exists j > 1 \text{ s.t. } \widehat{U} = \widehat{V}_j \right] \leq p^w \leq n^{-\alpha f(u)u}.$$

By Lemma 2.3 we know that there are at most $\frac{n(e\Delta)^u}{u}$ connected spanning subgraphs of cardinality u , thus we can bound the probability of appearance of a small connected component of cardinality u .

$$\Pr \left[\exists j > 1 \text{ s.t. } |\widehat{V}_j| = u \right] \leq \frac{n(e\Delta)^u}{u} \cdot n^{-\alpha f(u)u}. \quad (2)$$

Setting $\varepsilon = \min\{kf(k) - \frac{1}{\alpha} : k \geq K\}$, the definition of K implies ε is a positive constant. Applying (2) and summing over all possible values of u , we can bound the probability there will be in \widehat{G} a small connected component of cardinality at least K . First assume $K < \lceil \frac{2}{c\alpha} \rceil$.

$$\begin{aligned} \Pr \left[\exists j > 1 \text{ s.t. } |\widehat{V}_j| \geq K \right] &\leq \sum_{u=K}^{\bar{u}} \frac{n(e\Delta)^u}{u} \cdot n^{-\alpha f(u)u} \\ &\leq \sum_{u=K}^{\lceil \frac{2}{c\alpha} \rceil - 1} n^{1-\alpha u f(u) + o(1)} + \sum_{u=\lceil \frac{2}{c\alpha} \rceil}^{\bar{u}} n^{1-u(c\alpha - o(1))} \\ &\leq \left(\left\lceil \frac{2}{c\alpha} \right\rceil - K \right) n^{-\alpha \varepsilon + o(1)} + O \left(n^{2-\alpha - \lceil \frac{2}{c\alpha} \rceil (c\alpha - o(1))} \right) = o(1). \end{aligned}$$

Finally, if $K \geq \lceil \frac{2}{c\alpha} \rceil$ the above computation becomes simpler as we are left with only the second summand in the second line and the statement holds in that case as well. \square

Having shown that the small connected components of \widehat{G} are w.h.p. of bounded size, we move on to show that larger connected subsets of \widehat{G} w.h.p. expand.

Lemma 2.5. *W.h.p. every connected subset $U \subseteq \widehat{V}$ of \widehat{G} of cardinality u s.t. $K \leq u \leq \frac{\widehat{n}}{2}$ satisfies $|N_{\widehat{G}}(U)| \geq \frac{\alpha c}{4}u$.*

Proof. Similarly to the proof of Lemma 2.4 let $W = N_G(U)$ and $\widehat{W} = N_{\widehat{G}}(U)$ be the neighborhoods of U in G and \widehat{G} , respectively, and let w and \widehat{w} denote their respective cardinalities. By Lemma 2.4 we have that w.h.p. every connected subset of vertices U of cardinality $K \leq u \leq \frac{4}{\alpha c}$ is not disconnected from the graph, and thus has at least one edge leaving it. Setting $\eta = \frac{\alpha c}{4}$ this implies that for every such connected subset U , $\widehat{w} \geq \eta u$. Assuming $\frac{4}{\alpha c} < u \leq \frac{\widehat{n}}{2}$, relying on $\widehat{w} \sim \mathcal{B}(w, 1 - p)$ we have

$$\Pr[\widehat{w} < \eta u] \leq \binom{w}{\lfloor \eta u \rfloor} \cdot p^{w - \lfloor \eta u \rfloor} \leq \left(\frac{ew}{\eta u}\right)^{\eta u} p^{uf(u) - \eta u} \leq \left(\frac{e\Delta}{\eta}\right)^{\eta u} n^{-\alpha u(f(u) - \eta)}.$$

To bound the probability there exists a connected subset in \widehat{G} of cardinality u whose neighborhood contains less than ηu vertices, we apply the above with the union bound on all connected subsets of G from Lemma 2.3 as follows.

$$\begin{aligned} \frac{n(e\Delta)^u}{u} \cdot \left(\frac{e\Delta}{\eta}\right)^{\eta u} \cdot n^{-\alpha u(f(u) - \eta)} &\leq \eta^{-\eta u} \cdot (e\Delta)^{u(\eta+1)} \cdot n^{1 - \alpha u(f(u) - \eta)} \\ &\leq \left(\frac{(e\Delta)^{\eta+1}}{n^{\alpha c/4 \eta}}\right)^u \cdot n^{-1} = o(n^{-1}). \end{aligned}$$

The inequality from the first to the second line follows from the fact that $1 - \alpha u(f(u) - \frac{\alpha c}{4}) \leq -(1 + \frac{\alpha u c}{4})$, or equivalently $\alpha u f(u) - \frac{\alpha u c}{4}(1 + \alpha) \geq \frac{\alpha u c}{2} \geq 2$ using that $\alpha \leq 1$ and $u > \frac{4}{\alpha c}$. Summing over all possible values of u implies that w.h.p. there is no connected subset U in \widehat{G} of cardinality at least $\frac{4}{\alpha c}$ that satisfies $|N_{\widehat{G}}(U)| < \eta|U|$, completing the proof. \square

Lemma 2.5 states that w.h.p. all connected subsets of $G[\widehat{V}_1]$ expand. As G is of bounded maximum degree, this is sufficient to imply that w.h.p. all subsets of $G[\widehat{V}_1]$ expand.

Lemma 2.6. *W.h.p. $G[\widehat{V}_1]$ is a β -expander, where $\beta = \frac{1}{\Delta} \cdot \min\{\frac{1}{K}, \frac{\alpha c}{4}\}$.*

Proof. Set $\eta = \frac{\alpha c}{4}$ as defined in Lemma 2.5, and $\gamma = \min\{\frac{1}{K}, \eta\}$. For every $U \subseteq \widehat{V}_1$ of cardinality $|U| = u \leq K \leq 1/\gamma$, trivially $|N_{\widehat{G}}(U)| \geq 1 \geq \gamma u$, as U has at least one edge emitting out of it. Assume $u > K$ and denote by U_1, \dots, U_t the decomposition of U to its connected subsets, and by u_1, \dots, u_t their respective cardinalities. As every connected subset U_i satisfies w.h.p. $|N_{\widehat{G}}(U_i)| \geq \gamma u_i$ by Lemma 2.5, it follows that w.h.p. $|N_{\widehat{G}}(U)| \geq \frac{\gamma}{\Delta}u$ completing the proof. \square

Combining Lemmata 2.2, 2.4, and 2.6 completes the proof of Theorem 1.2

3 Applications to different expander graph families

3.1 Expansion via the spectrum of a graph

The *adjacency matrix* of a graph G on n vertices labeled by $\{1, \dots, n\}$, is the $n \times n$ binary matrix, $A(G)$, where $A(G)_{ij} = 1$ iff $i \sim j$. The *combinatorial Laplacian* of G is the $n \times n$ matrix $L(G) = D - A(G)$ where D

is the diagonal matrix defined by $D_{i,i} = d_G(i)$. It is well known that for every graph G , the matrix $L(G)$ is positive semi-definite (see e.g. [6]), and hence has an orthonormal basis of eigenvectors and all its eigenvalues are non-negative. We denote the eigenvalues of $L(G)$ in the ascending order by $0 = \sigma_0 \leq \sigma_1 \leq \dots \leq \sigma_{n-1}$, where σ_0 corresponds to the eigenvector of all ones. We denote by $\tilde{d} = \tilde{d}(G)$ the average degree of G , and let $\theta = \theta(G) = \max\{|\tilde{d} - \sigma_i| : i > 0\}$. The celebrated expander mixing lemma (see e.g. [3]) and its generalization to the non-regular case (see e.g. [6]) state roughly that the smaller θ is, the more random-like is the graph. This easily implies several corollaries on the distribution of edges in the graph. In particular, one can deduce the following expansion property of G in terms of \tilde{d} and θ . Its full proof can be found in [6].

Proposition 3.1. *Let G be an graph on n vertices. Then G is an $h_{n,\tilde{d},\theta}$ -expander, where*

$$h_{n,\tilde{d},\theta}(i) = \frac{\tilde{d}^2 - \theta^2}{\theta^2 + \tilde{d}^2 \frac{i}{n-i}} \quad \text{for } 1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor. \quad (3)$$

Assume G is a graph of bounded maximum degree, implying $\tilde{d} = O(1)$, and let $H(i) = i \cdot h_{n,\tilde{d},\theta}(i)$. Straightforward analysis implies $H(i)$ is monotonically increasing for $1 \leq i \leq \left\lfloor \frac{n\theta}{\tilde{d} + \theta} \right\rfloor$, and monotonically decreasing for $\left\lfloor \frac{n\theta}{\tilde{d} + \theta} \right\rfloor \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor$. When $\tilde{d} - \theta > \varepsilon$ for some $\varepsilon > 0$, we have that $h_{n,\tilde{d},\theta} \geq c$, where $c = \frac{\tilde{d}^2 - \theta^2}{\tilde{d}^2 + \theta^2}$ is a constant depending on \tilde{d} and θ . We note that $H(\left\lfloor \frac{n}{2} \right\rfloor) = O(n) \gg 1/\alpha$. Setting $k = \frac{\theta^2}{(\tilde{d}^2 - \theta^2)\alpha} + 1$, we have that $k \leq \left\lfloor \frac{n\theta}{\tilde{d} + \theta} \right\rfloor$ and

$$H(k) = \left(\frac{\theta^2}{(\tilde{d}^2 - \theta^2)\alpha} + 1 \right) \cdot \left(\frac{\tilde{d}^2 - \theta^2}{\theta^2 + o(1)} \right) > \frac{1}{\alpha}.$$

Our analysis of $H(i)$ implies that the value K defined in (1) satisfies $K \leq k$. Proposition 3.1 thus enables us to apply Theorem 1.2 to such graphs.

Theorem 3.2. *For every fixed $\alpha, \varepsilon > 0$ and fixed $\Delta \geq 0$ there exists a constant $\beta > 0$, such that if G is a graph on n vertices of maximum degree Δ , and $\tilde{d} - \theta > \varepsilon$, then w.h.p. \hat{G} has a connected component of size $n - o(n)$ that is a β -expander, and all other components are of cardinality at most $\frac{\theta^2}{(\tilde{d}^2 - \theta^2)\alpha}$.*

3.2 (n, d, λ) -graphs

When the graph G is d -regular, $L(G) = dI - A(G)$, and hence if $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1}$ is the spectrum of $A(G)$ we have that $\lambda_i = d - \sigma_i$. In the case of d -regular graphs it is customary and, arguably, more natural to use the spectrum of $A(G)$ rather than of $L(G)$ to address expansion properties of the graph. As the largest eigenvalue of A is clearly $\lambda_0 = d$ and it is maximal in absolute value, we have that $\theta(G) = \max\{|\lambda_1(G)|, |\lambda_{n-1}(G)|\}$. In the case of d -regular graphs it is customary to denote $\theta(G)$ by $\lambda(G) = \lambda$, and to call such a graph G an (n, d, λ) -graph. For an extensive survey of fascinating properties of (n, d, λ) -graphs the reader is referred to [12].

In the case of (n, d, λ) -graphs Proposition 3.1 and Theorem 3.2 translate to the following.

Proposition 3.3. *Let G be an (n, d, λ) -graph, then G is an $h_{n,d,\lambda}$ -expander, where*

$$h_{n,d,\lambda}(1) = d; \quad \text{and} \quad h_{n,d,\lambda}(i) = \frac{d^2 - \lambda^2}{\lambda^2 + d^2 \frac{i}{n-i}} \quad \text{for } 2 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor. \quad (4)$$

Theorem 3.4. *For every fixed $\alpha, \varepsilon > 0$ and fixed $d \geq 3$ there exists a constant $\beta > 0$, such that if G is an (n, d, λ) -graph where $d - \lambda > \varepsilon$, then w.h.p. \hat{G} has a connected component of size $n - o(n)$ that is a β -expander, and all other components are of cardinality at most $\frac{\lambda^2}{(d^2 - \lambda^2)\alpha}$.*

The next two propositions allow us to get improved bounds on the sizes of the small connected components of \hat{G} . In Proposition 3.5 we are interested in the values of α for which \hat{G} is w.h.p. connected, and in Proposition 3.6 in the values for which w.h.p. the small connected components of \hat{G} are all isolated vertices. We compute these values of α under some additional assumptions on the (n, d, λ) -graph. Specifically, we require the graph to be locally “sparse” and the *spectral gap*, i.e. $d - \lambda$, to be relatively large. Although these constraints may seem somewhat artificial, they arise naturally in the setting of random d -regular graphs as will be exposed in Section 3.3. For any graph $G = (V, E)$ we denote by

$$\rho(G, M) = \max \left\{ \frac{e(U)}{|U|} : U \subseteq V \text{ s.t. } |U| \leq M \right\},$$

where $e(U)$ denotes the number of edges of G that have both endpoints in U .

Proposition 3.5. *For every $\alpha > \frac{1}{d}$ and fixed $d \geq 3$, if G is an (n, d, λ) -graph satisfying $\lambda \leq 2\sqrt{d-1} + \frac{1}{40}$ and $\rho(G, d+29) \leq 1$, then w.h.p. \hat{G} is connected.*

Proof. We prove that such a graph G is an f -expander where $if(i) \geq d$ for every $1 \leq i \leq \lfloor \frac{n}{2} \rfloor$. Lemma 2.4 will then imply that $K = 1$, and hence \hat{G} is w.h.p. connected. Proposition 3.3 guarantees that $f(i) \geq h_{n,d,\lambda}(i)$ (with $h_{n,d,\lambda}$ as defined in (4)). Taking $i \geq 30$ and plugging our assumption on λ in the definition of $h_{n,d,\lambda}$, we have that for every $d \geq 3$

$$f(30) \geq \frac{d^2 - 4(d-1) - \frac{\sqrt{d-1}}{10} - \frac{1}{1600}}{4(d-1) + \frac{\sqrt{d-1}}{10} + \frac{1}{1600} + d^2 \frac{30}{n-30}} > \frac{d}{30}.$$

The analysis of H therefore guarantees that $if(i) \geq d$ for all $i \geq 30$. Now, let U be a subset of vertices of cardinality $u \leq 29$, and set $s = |N_G(U)|$ and $w = |U \cup N_G(U)|$. It now suffices to show that $s \geq d$ for such a set U . If $u = 1$, trivially $s = d$, as every vertex has d neighbors. Now, our assumption on G implies that all triangles in the graph must be edge disjoint. Taking $u = 2$, if the two vertices in U are non-adjacent trivially $s \geq d$, and if they are adjacent, they must have at most one common neighbor, implying $s \geq 2d - 3 \geq d$. Taking $3 \leq u \leq 29$, if $w \geq u + d$ we are done. Otherwise, the assumption on G implies $e(U) \leq u$ and $e(U \cup N_G(U)) \leq w$, and hence $du - u \leq du - e(U) \leq e(U \cup N_G(U)) \leq w = u + s$. This implies $s \geq u(d-2) \geq d$ which completes the proof. \square

For any graph G we denote by $t(G)$ the number of triangles in G . To analyze the values of α for which w.h.p. all small connected components of \hat{G} are isolated vertices, we additionally require that the number of triangles in G is bounded by a certain positive power of n . This requirement as well is quite natural in the case of random d -regular graphs.

Proposition 3.6. *For every $\alpha > \frac{1}{2(d-1)}$ and fixed $d \geq 3$, if G is an (n, d, λ) -graph satisfying $\lambda \leq 2\sqrt{d-1} + \frac{1}{40}$, $\rho(G, 39 + 2(d-1)) \leq 1$, and $t(G) = O(n^{\frac{2d-3}{2(d-1)}})$, then w.h.p. all small connected components of \hat{G} are isolated vertices.*

Proof. Following the spirit of the proof of Proposition 3.5, we would like to show that G is an f -expander where $if(i) \geq 2(d-1)$ for every $2 \leq i \leq \lfloor \frac{n}{2} \rfloor$, which completes the proof by using Lemma 2.4. By Proposition

3.3 we can assume that $f(i) \geq h_{n,d,\lambda}(i)$. Our assumption on λ gives

$$f(40) \geq \frac{d^2 - 4(d-1) - \frac{\sqrt{d-1}}{10} - \frac{1}{1600}}{4(d-1) + \frac{\sqrt{d-1}}{10} + \frac{1}{1600} + d^2 \frac{40}{n-40}} > \frac{d-1}{20},$$

which in turn, using our analysis of H , implies $if(i) \geq 2(d-1)$ for $i \geq 40$. When trying to complete the proof by showing that $f(i) \geq \frac{2d-1}{i}$ for $2 \leq i \leq 39$, it turns out that in this setting this will not be the case, as there can be small subsets that violate this strict expansion requirement. Fortunately, we can prove that there cannot be too many such subsets, which allows us to prove the above probabilistic statement.

Let U be a subset of vertices of cardinality $2 \leq u \leq 39$, and set $s = |N_G(U)|$ and $w = |U \cup N_G(U)|$. We call U *exceptional* if $s < 2(d-1)$. Let x_i denote the number of exceptional sets of cardinality i .

If $4 \leq u \leq 39$, our assumption on G implies $e(U) \leq u$ and $e(U \cup N_G(U)) \leq w$. It follows that $du - u \leq e(U \cup N_G(U)) \leq w = u + s$, implying $s \geq u(d-2) \geq 2(d-1)$. If $u = 2$ and the two vertices of U are non-adjacent then $2d \leq e(U \cup N_G(U)) \leq w = 2 + s$, hence $s \geq 2(d-1)$. If the two vertices are adjacent, but are not part of a triangle, then again $s \geq 2(d-1)$. For $u = 3$, if U spans at most one edge, then $3d - 1 \leq e(U \cup N_G(U)) \leq w = 3 + s$, yielding $s \geq 2(d-1)$. If U spans a triangle and $d \geq 4$, as all triangles of G must be edge disjoint we get that $s \geq 3(d-2) \geq 2(d-1)$. If U spans exactly two edges, easy case analysis, relying on the fact that no small subgraph spans more edges than vertices, shows that $s \geq 3d - 5 \geq 2(d-1)$.

Lemma 2.4 and the previous computation assure that for $d \geq 4$ w.h.p. all small connected components have at most two vertices, and for $d = 3$, w.h.p. all small connected components have at most three vertices.

We conclude by showing that since in both cases there are only a small number of exceptional sets, w.h.p. all small connected components will be isolated vertices. Similarly to the proof of Lemma 2.4, we bound the probability of appearance of a connected component of cardinality 2. The exceptional sets of cardinality 2 are edges that participate in a triangle and all triangles in G are edge disjoint, therefore there are $x_2 = 3t(G)$ such exceptional sets, and each has exactly $2d-3$ neighbors. Going over all connected sets of G of cardinality 2, i.e. the edges of G , we bound the probability that one of these sets becomes disconnected.

$$\begin{aligned} \Pr \left[\exists j > 1 \text{ s.t. } |\widehat{V}_j| = 2 \right] &\leq x_2 p^{2d-3} + \left(\frac{dn}{2} - x_2 \right) p^{2(d-1)} \\ &\leq O \left(n^{\frac{2d-3}{2(d-1)} - \alpha(2d-3)} \right) + O \left(n^{1-2\alpha(d-1)} \right) = o(1). \end{aligned}$$

The above completes the proof when $d \geq 4$. We are left with the case of exceptional triples that may exist when $d = 3$. Since the exceptional sets of cardinality 3 are the triangles in G , there are exactly $x_3 = t(G)$ such exceptional sets each having exactly 3 neighbors. Very similarly to the preceding computation, we go over all connected sets of G of cardinality 3, i.e. sets that span two or three edges, and compute the probability that one of these sets become disconnected. Recall that for $d = 3$ we have that $\alpha > \frac{1}{4}$.

$$\begin{aligned} \Pr \left[\exists j > 1 \text{ s.t. } |\widehat{V}_j| = 3 \right] &\leq x_3 p^3 + (3n - 3t(G)) p^4 \\ &\leq O \left(n^{\frac{3}{4} - 3\alpha} \right) + O \left(n^{1-4\alpha} \right) = o(1). \end{aligned}$$

□

Propositions 3.5 and 3.6 are easily seen to be optimal in some sense, for if $\alpha \leq \frac{1}{d}$ or $\alpha \leq \frac{1}{2(d-1)}$, then the expected number of isolated vertices or edges respectively is greater than 1.

3.3 Random d -regular graphs

Consider the random graph model consisting of the uniform distribution on all d -regular graphs on n vertices (where dn is even), and denote this probability space by $\mathcal{G}_{n,d}$. Assume throughout this section that $d \geq 3$ is a constant. Let G be a graph sampled from $\mathcal{G}_{n,d}$. Note that the multiplicity of the eigenvalue d of the graph G is w.h.p. 1 as G is w.h.p. connected and non-bipartite (see e.g. [15]), hence w.h.p. $\lambda(G) < d$. Friedman, confirming a conjecture of Alon, gives an accurate evaluation of $\lambda(G)$ for most random d -regular graphs when d is a constant.

Theorem 3.7 (Friedman [7]). *For any $\varepsilon > 0$ and fixed $d \geq 3$, if G is sampled from $\mathcal{G}_{n,d}$ then w.h.p.*

$$\lambda(G) \leq 2\sqrt{d-1} + \varepsilon. \quad (5)$$

Combining Theorems 3.4 and 3.7, implies explicitly the first part of Theorem 1.1.

Corollary 3.8. *For every fixed $\alpha > 0$ and fixed $d \geq 3$ there exists a constant $\beta > 0$, such that if G is a graph sampled from $\mathcal{G}_{n,d}$, then w.h.p. \hat{G} has a connected component of size $n - o(n)$ that is a β -expander and all other components are of cardinality at most $\frac{4(d-1)}{\alpha(d-2)^2} + 1$.*

The second part of Theorem 1.1 analyzes the values of α for which w.h.p. the graph \hat{G} is connected, and the values of α for which the w.h.p. small connected components are all isolated vertices. Plugging Theorem 3.7 into Theorem 3.4, as above, implies a similar result, but not as strong.

To get Theorem 1.1 in full, and even to improve it, we use Propositions 3.5 and 3.6. To do so, we state the following well known asymptotic properties of $\mathcal{G}_{n,d}$ (see e.g. [15]). Let G be a graph sampled from $\mathcal{G}_{n,d}$, for any fixed $d \geq 3$, then w.h.p. the minimal distance between two cycles of constant length in G is $\omega(1)$. This statement is equivalent to saying that for every constant $M > 1$ w.h.p. $\rho(\mathcal{G}_{n,d}, M) \leq 1$. Moreover, as n tends to infinity $t(\mathcal{G}_{n,d}) \sim \text{Poisson}\left(\frac{(d-1)^3}{6}\right)$, and by so Markov's inequality w.h.p. $t(\mathcal{G}_{n,d}) = O(n^{\frac{2d-3}{2(d-1)}})$ (with room to spare). Now, using Propositions 3.5 and 3.6 combined with Corollary 3.8 we get the desired result for $\mathcal{G}_{n,d}$.

Theorem 3.9. *For every fixed $\alpha > 0$ and $d \geq 3$ there exists a constant $\beta > 0$, such that if $p = n^{-\alpha}$ and G is a graph sampled from $\mathcal{G}_{n,d}$, then w.h.p. \hat{G} has a connected component of size $n - o(n)$ that is a β -expander and all other components are of cardinality at most $\frac{4(d-1)}{\alpha(d-2)^2} + 1$. Moreover,*

1. *if $\alpha > \frac{1}{2(d-1)}$, w.h.p. all small connected components of \hat{G} are isolated vertices.*
2. *if $\alpha > \frac{1}{d}$, w.h.p. \hat{G} is connected.*

It should be noted that Theorem 3.9 improves upon Theorem 1.1 for the values of α guaranteeing that \hat{G} stays connected w.h.p.. As mentioned in the Section 3.2, this improvement is best possible, for if $\alpha \leq \frac{1}{d}$, then the expected number of isolated vertices will be at least one, and by some standard concentration arguments it can also be shown that the number of isolated vertices is highly concentrated around this expectation. Hence, for $\alpha \leq \frac{1}{d}$ the graph \hat{G} has isolated vertices, and is thus disconnected, with some probability bounded away from 0. As a final note, it should be mentioned that in the original statement of the main result of [9], it is proved that w.h.p. all small connected components are trees, and that for $\alpha > \frac{1}{2(d-1)}$ w.h.p. the number of isolated vertices is $o(n^{(d-2)/2(d-1)})$. These results as well can be derived from simple probabilistic arguments based on properties of $\mathcal{G}_{n,d}$, but we omit these technical details.

4 Unbounded expansion of small sets

So far we have considered graphs of bounded maximum degree (and in particular d -regular graphs for $d = O(1)$) that expand by a constant factor. When considering graphs that expand sets of sub-linear cardinality by an $\omega(1)$ factor (in particular in such graphs $\delta(G) = \omega(1)$, i.e. the minimal degree of G goes to infinity with n) a simple union bound argument implies the following result. The proof is quite similar to those we have previously presented, only in this case we can use a union bound over all subsets of vertices with no need to go over all connected subsets first, i.e. we do not make use of Lemma 2.3 .

Theorem 4.1. *For every fixed $\alpha, c, \varepsilon > 0$ if G is an f -expander graph on n vertices where $f(u) = \omega(1)$ for every $u = o(n)$, and $f \geq c$, then w.h.p. \widehat{G} is a $(c - \varepsilon)$ -expander.*

Proof. Let $U \subseteq V$ be a subset of vertices of cardinality $u \leq \frac{n}{2}$, and let $W = N_G(U)$ be its neighborhood in G , where $|W| = w$. Set $\beta = c - \varepsilon$, and let us denote a subset of vertices U as *bad* if $\widehat{U} = U$ and $\widehat{w} < \beta u$. If \widehat{G} is not a β -expander then it must contain such a bad set. We bound the probability of a subset U to be bad by

$$\Pr[U \text{ is bad}] \leq \binom{w}{\lfloor \beta u \rfloor} \cdot p^{w - \lfloor \beta u \rfloor} \leq \left(\frac{ew}{\beta u} \right)^{\beta u} p^{u(f(u) - \beta)}.$$

Assuming $u = o(n)$, we have

$$\Pr[\exists U \subseteq V \text{ s.t. } |U| = u \text{ and } U \text{ is bad}] \leq \binom{n}{u} \left(\frac{ew}{\beta u} \right)^{\beta u} p^{u(f(u) - \beta)} \leq n^{u(1 + \beta(1 + \alpha) + o(1) - \alpha\omega(1))} = o(n^{-1}).$$

In the case that $\Theta(n) = u \leq \frac{n}{2}$, we have

$$\Pr[\exists U \subseteq V \text{ s.t. } |U| = u \text{ and } U \text{ is bad}] \leq \binom{n}{u} \left(\frac{ew}{\beta u} \right)^{\beta u} p^{u(f(u) - \beta)} \leq n^{u(o(1) + o(1) - \alpha(c - \beta))} = o(n^{-1}).$$

Applying the union bound over all possible values of u completes the proof. \square

It should be noted that Theorem 4.1 implies that when $p = n^{-\alpha}$ for any fixed $\alpha > 0$, \widehat{G} is w.h.p. an expander, and in particular stays connected as opposed to the case of bounded maximum degree.

When $d = o(\sqrt{n})$, Broder et al. [5, Lemma 18] provide an upper bound on the second eigenvalue of most of the d -regular graphs.

Theorem 4.2 (Broder et al. [5]). *For $d = o(\sqrt{n})$, if G is sampled from $\mathcal{G}_{n,d}$ then w.h.p.*

$$\lambda(G) = O(\sqrt{d}). \tag{6}$$

Plugging Theorem 4.2 into Proposition 3.3 assures that w.h.p. all conditions needed in Theorem 4.1 are met when the graph sampled from $\mathcal{G}_{n,d}$ for $1 \ll d \ll \sqrt{n}$, and hence we get the following result.

Theorem 4.3. *For every fixed $\alpha > 0$ and $1 \ll d \ll \sqrt{n}$ there exists a constant $\beta > 0$, such that if G is a graph sampled from $\mathcal{G}_{n,d}$, then w.h.p. \widehat{G} is a β -expander.*

When sampling a graph from the binomial random graph model $\mathcal{G}_{n,p}$ (i.e. the probability space of all graphs on n labeled vertices, where each pair of vertices is chosen to be an edge independently with probability p) with $p = \frac{d}{n}$ for $d = \Omega(\sqrt{n})$, the graph is easily seen to be “almost d -regular” as all degrees

of the vertices are highly concentrated around d . Furthermore, it can be easily shown that when the initial graph is sampled from $\mathcal{G}_{n,p}$ with the prescribed values of p , a similar claim to Theorem 4.1 holds. Therefore, one should expect Theorem 4.1 to extend to values of $d = \Omega(\sqrt{n})$, but unfortunately, the techniques that are commonly used to deal with random regular graphs seem to fail for these higher values of d .

We note that in [4] the authors prove a result on the distribution of edges in $\mathcal{G}_{n,d}$ for $d = o(\sqrt{n})$, that can be easily used to derive vertex-expansion properties of $\mathcal{G}_{n,d}$ for $1 \ll d \ll \sqrt{n}$, and combined with Theorem 4.1 provides an alternative proof of Theorem 4.3.

5 Concluding remarks and open problems

In this paper we analyzed the process of deleting uniformly at random vertices from an expander graph. We have shown that for small enough deletion probabilities the resulting graph w.h.p. retains some expansion properties (if not in the graph itself then in its largest connected component). We have also proved that for these deletion probabilities w.h.p. all small connected components must be of bounded size. Lastly, we have shown how this result can be applied to the random d -regular graph model for $d = o(\sqrt{n})$.

In Section 3.3, in order to apply our results from previous sections to the case of random d -regular graphs, we made use of several theorems that describe some properties that occur w.h.p. in graphs that are sampled from $\mathcal{G}_{n,d}$, such as Theorem 3.7 of Friedman [7]. This very strong result, whose proof is far from simple, seems to be an overkill to prove our claims. One could go about by showing that graphs from $\mathcal{G}_{n,d}$ w.h.p. possess some expansion property (by analyzing the model directly using, e.g., the Configuration Model or the Switching Technique) and then by applying Theorem 1.2 directly. This method would undoubtedly provide a proof that does not require any “heavy duty machinery”, but does require more meticulous computations. Nonetheless, we hope that the reader finds the use of the connection between spectral graph theory and expansion properties (or pseudo-randomness of a graph) to be both elegant and concise.

In light of Theorem 4.3 it would be interesting to analyze the expansion properties of random d -regular graphs for $d = \omega(1)$ for higher values of p , i.e. taking $p = n^{-o(1)}$, as for $d = \omega(1)$ it is no longer true trivially that for these values of p w.h.p. there will be long induced paths in \widehat{G} .

Acknowledgements

The authors would like to thank the anonymous referees for their helpful corrections and comments. The first author would like to thank Itai Benjamini for discussing this problem with him.

References

- [1] N. Alon, I. Benjamini and A. Stacey, Percolation on finite graphs and isoperimetric inequalities, *Annals of Probability*, Vol. 32(3):1727–1745, 2004.
- [2] N. Alon, H. Kaplan, M. Krivelevich, D. Malkhi and J. Stern, Scalable secure storage when half the system is faulty, *Information and Computation*, Vol. 174(2):203–213, 2002.

- [3] N. Alon and J. Spencer, **The Probabilistic Method**, 2nd ed., Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, 2000.
- [4] S. Ben-Shimon and M. Krivelevich, Random regular graphs of non-constant degree: edge distribution and applications, preprint. <http://arxiv.org/abs/math/0511343>.
- [5] A. Broder, A. Frieze, S. Suen, and E. Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM Journal on Computing*, Vol. 28(2):541–573, 1999.
- [6] F. Chung, Discrete isoperimetric inequalities, *Surveys in Differential Geometry: Eigenvalues of Laplacians and other geometric operators*, International Press, (A. Grigor’yan and S. T. Yau, Eds.) Vol. IX:53–82, 2004.
- [7] J. Friedman, *A Proof of Alon’s Second Eigenvalue Conjecture and Related Problems*, Memoirs of the AMS, to appear.
- [8] A. Frieze, M. Krivelevich and R. Martin, The emergence of a giant component in random subgraphs of pseudo-random graphs, *Random Structures and Algorithms*, Vol. 24(1):42–50, 2004.
- [9] C. Greenhill, F. B. Holt and N. C. Wormald, Expansion properties of a random regular graph after random vertex deletions, *European Journal of Combinatorics*, Vol. 29(5):1139–1150, 2008.
- [10] F. B. Holt, V. Bourassa, A. M. Bosnjakovic and J. Popovic, SWAN - Highly reliable and efficient networks of true peers, *CRC Handbook of Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks*, (J. Wu Ed.), 787–811, 2005.
- [11] S. Hoory, N. Linial and A. Wigderson, Expander graphs and their applications. *Bulletin of the AMS*, Vol. 43(4):439–561, 2006.
- [12] M. Krivelevich and B. Sudakov, Pseudo-random graphs, *More sets, graphs and numbers*, *Bolyai Society Mathematical Studies*, (E. Györi, G. Katona and L. Lovász, Eds.), Vol. 15:199–262, 2006.
- [13] D. Knuth, **The Art of Computer Programming**, Vol. I., 1969.
- [14] A. Nachmias, Mean-field conditions for percolation on finite graphs, preprint. <http://www.arxiv.org/abs/0709.1719>.
- [15] N. C. Wormald, Models of random regular graphs, *Surveys in Combinatorics, London Mathematical Society Lecture Note Series*, (J. Lamb and D. Preece, Eds.), Vol. 276:239–298, 1999.